

Comments to Jean-Claude Burgelman's article *Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story)*

Peter Wittenburg[†]

Max Planck Computing and Data Facility, Gießenbachstraße 2, 85748 Garching, Germany

Citation: Wittenburg, P.: Comments on Jean-Claude Burgelman's article *Politics and Open Science: How the European Open Science Cloud became reality (the untold story)*. Data Intelligence 3(1), 47-51 (2021). doi: 10.1162/dint_a_00076

1. PREFACE

Coming from an institute that was devoted to analysing data streams of different sorts from its beginning to understand how the human brain is processing language and how language is supporting cognition, building efficient data infrastructures of different scope was a key to research excellence. While first local infrastructures were sufficient, it became apparent in the 90s that local data would not be sufficient anymore to satisfy all research needs. It was a logical step to first take responsibilities in setting up the specific DOBES (Dokumentation bedrohter Sprachen) infrastructure focussing on languages of the world, then the community-wide CLARIN RI (European Research Infrastructure for Language Resources and Technology) and later the cross-disciplinary EUDAT data infrastructure [1,2,3]. Realising the huge heterogeneity in data practices, it was also a logical step to start the Research Data Alliance (RDA) [4] as a truly bottom-up initiative to discuss harmonisation across disciplines and across borders. On this background, determined by always looking for concrete results, the European Open Science Cloud (EOSC) process had Kafka-esc characteristics to me, despite the many interactions I had with EOSC key persons and other colleagues involved. Talking at a level where the technological core remained widely absent was difficult to do for me. Due to Jean-Claude Burgelman's (JCB) excellent paper I finally understood that excluding the discussions about the core was the only chance to get EOSC accepted. Of course, the discussions about the EOSC core would have to happen at a certain moment and obviously eternal types of disputes would determine these discussions. Therefore, the fallback on the analogy with Greek tragedies was an excellent idea by JCB.

[†] Corresponding author: Peter Wittenburg (Email: peter.wittenburg@mpcdf.mpg.de; ORCID: 0000-0003-3538-0106).

2. EOSC IN BIZARRE TIMES

EOSC has grown in bizarre times in so far as we are all using the global web as a primary medium for exchanging information on the one hand, however, recognising that the architecture for global data management needs to be changed on the other hand. Global players started the Data Transfer Project [5] indicating that even big companies are looking for alternatives, yet without admitting that metadata and identification will be key to success. Even T. Berners-Lee stated recently that “there is a feeling, a zeitgeist, that change is really overdue” [6]. Some believe that FAIR Digital Objects [7,8] are the way out to solve some critical core aspects of the evolving global data infrastructure, while others are preferring a smooth process of upgrading services or believe that we should leave solution finding to the big companies, since they have the money to enforce a big change.

In parallel, we have engaged discussions in all sectors about the questions associated with data sovereignty: who has access to data, who will be able to sell services on data, will we in using data become dependent on monopolies again? EOSC was a natural policy response in Europe to these questions. It is also a policy response to other imminent questions in Data Science (DS): will we end up in Digital Dark Times, will DS lack all kinds of responsibility and accountability? We also realise that large communities such as the librarians that played such a great role in classical science are looking to find their place in this new digital data landscape and that countries want to claim leadership in specific areas of expertise and services. The landscape is highly complex and competitive, and until now researchers did not really play a primary role in these debates. In fact, deep insights indicate that researchers in the labs did not care yet about EOSC [9].

3. EOSC AS POLICY ACHIEVEMENT

To me having EOSC is indeed a great policy achievement for Europe and, for everyone visible, it put global data management as a priority on the agenda. As JCB admits, it is yet an empty box. While I was always trying to convince policy players to start discussing the technological questions at an early point in time, I now realised after having participated in some EOSC discussions, that due to the differences in approaches and interests it was indispensable for the key actors to leave the technological core in the dark as long as possible. After having read JCB's paper, it is the first time that I understood that Europe must be happy now to have a policy instrument, although it does not mean that Europe is ready to make proper advances in future. Due to decisions about the governance structure the “Brussels bubble” is now replaced by a bubble in which member state (MS) delegates and existing initiatives widely define the rules of the game. Delegates of these stakeholders often seem to mix short-term interests with arguments about technological strategies, making it hard to believe that wise decisions can be taken.

The “EOSC emptiness” for me includes two primary and related questions: (1) What is the ambition behind all our investments, i.e. are we targeting for 5, 10, 50 or 100+ years? (2) Do we believe that smooth evolution of existing services will provide the solution for the huge challenges or are we ready to dare thinking about disruptive steps which implies more risks? EC and member states (MSs) are investing so huge

sums in the coming decade that it is important to give clear answers to these questions. With the current vagueness of the goals, everyone can sell anything what is being done under the label EOSC, even scientifically useless portals supported by millions of Euro can be sold as EOSC contributions. This will not help Europe taking a leading role.

4. EOSC AND RESEARCHERS

George Strawn hints to one of the major paradoxes that we are faced with, when he argues that “standards are good for science, but bad for the scientists” [10]. Deep insights show that most researchers in the labs like to continue with the methods they are used to and only like to change, if they see clear short-term benefits. This implies that researchers, in general, will not push innovative infrastructure work which comes along with new standards, regulations and eventually disruptions. They were not interested in technological innovations such as TCP/IP despite its revolutionary character, they only became interested when new tools became available enriching the set of options for their research—an attitude which is acceptable given the pressure to show results. Until now EOSC discussions were far away from the researchers in the labs. EOSC was mainly discussed in a bubble of librarians, archivists, research infrastructure and e-Infrastructure officials being ready to sign declarations motivated in many cases by rather opportunistic attitudes. We should not interpret this as a critique since it is a real challenge to organise a broad discussion process about initiatives such as EOSC. And here I would like to slightly disagree with JCB: to me, EOSC until now is indeed an invention from Brussels if we include all stakeholders dependent on funding streams from EC. Since we need to accept that big innovative steps in infrastructure technology were, in general, pushed ahead by others than the mass of researchers, we need to rely on voices that take a strategic view combined with insights in technological trends. Therefore, I appreciate the driving role of the EC until now.

5. EOSC WAY OUT

Now, EOSC is a fact in European policy, but only wise decisions will make it a success and prevent another Greek tragedy. Starting the discussions about the EOSC core is not too late, since finding decisions about the core of global data infrastructures will take more time than I hoped when starting RDA, for example. Also, the FAIR principles were a huge step, but we start realising that they are not sufficient to serve as the solution.

EOSC will have to overcome some of its weaknesses to be successful. (1) Although the creation of large infrastructures is of national interests, EOSC needs to find a way to separate technological discussions from political interests. (2) It needs to identify its ambitions—if all the funds will be spent to make small steps, it will not succeed, and decisions will be overruled by the next technology wave. Nevertheless, relying on a federated approach was a wise decision, since so much money has already been spent on building important components such as repositories. (3) Different approaches have been suggested, and their potential needs to be investigated by investing reasonable sums in test beds and reference architectures and taking risks. Politics needs to have experts who explored the landscape of solutions. Having used the cloud

metaphor in EOSC was probably a big error, since too many of my colleagues took it literally. Investing in cloud services, however, makes sense to make Europe independent with respect to large digital stores, but cloud technology is now very much known and does not address the core issues to be solved on the way towards an integrated and interoperable data domain (I2D2). (4) Given a decision about the scope of ambitions, EOSC needs to be able to organise a discussion process with widely independent experts to identify trends and evaluate the possible development of technologies.

JCB refers to the GAIA-X project [11], but this initiative was also established as an empty box without conceptual views. It seems that also European industry does not have a clear view about the key challenges for establishing the I2D2 which we will need in a few decades.

Finally, after having observed the EOSC process my question is whether Europe will be in the state of characterising the I2D2 domain and thus infer directions for investments that will be ground-breaking. We did not understand the core of the information wave and seem to not be able to organise a discussion process allowing us to anticipate the data wave and come to conclusions which is another paradox given the fact that European colleagues were and are pushing the agendas in RDA and around FAIR.

REFERENCES

- [1] DOBES. Available at: <https://dobes.mpi.nl/> Accessed 20 December 2020
- [2] CLARIN. Available at: <https://www.clarin.eu/> Accessed 20 December 2020
- [3] EUDAT. Available at: <https://www.eudat.eu/> Accessed 20 December 2020
- [4] RDA. Available at: <https://rd-alliance.org/> Accessed 20 December 2020
- [5] DTP. Available at: <https://datatransferproject.dev/> Accessed 20 December 2020
- [6] Tim Berners-Lee. Available at: <https://www.ft.com/content/01480644-3ca3-486e-907d-4abf8aac1719> Accessed 20 December 2020
- [7] Digital Objects. Available at: <http://hdl.handle.net/11304/5d760a3e-991d-11e5-9bb4-2b0aad496318> Accessed 20 December 2020
- [8] FAIR DO. Available at: <https://github.com/GEDE-RDA-Europe/GEDE/tree/master/FAIR%20Digital%20Objects> Accessed 20 December 2020
- [9] Jeffery, K., et al.: Not Ready for Convergence in Data Infrastructures. *Data Intelligence* 3(1), 116–135 (2021)
- [10] Strawn, G.: Open Science and the Hype Cycle. *Data Intelligence* 3(1), 43–46 (2021)
- [11] GAIA-X. Available at: <https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html> Accessed 20 December 2020

AUTHOR BIOGRAPHY



Peter Wittenburg was Executive Director of Research Data Alliance (RDA) Europe, Member of RDA Technical Advisory Board, Scientific Coordinator of European Data Infrastructure (EUDAT) and Technical Director of the CLARIN and DOBES Research Infrastructures. He set up and led the Technical Group with about 35 experts at Max Planck Institute (MPI) for Psycholinguistics and then led the Language Archiving Group with about 25 experts. Since 2000 he has played leading roles in a variety of European (funded by the European Commission) and national projects (funded by MPS, DFG, BMBF, NWO) and ISO initiatives (ISO TC37/SC4). He won the Heinz Billing Award of the Max Planck Society (MPS) for the advancement of scientific computation in 2011 and received an honorary doctorate from University Tübingen in 2013. ORCID: 0000-0003-3538-0106